

# Zhaoyi (Joey) Hou

website: joeyhou.github.io      email: joey.hou@pitt.edu

**Research Interest** Artificial Intelligence, Natural Language Processing, Commonsense Reasoning

**Education**

<b>University of Pittsburgh</b> Computer Science, Doctor of Philosophy Research Advisor: Dr. Xiang (Lorraine) Li	2023 - Present
<b>University of Pennsylvania</b> Data Science, Master of Science in Engineering Research Advisor: Dr. Chris Callison-Burch	2021 - 2023
<b>University of California, San Diego</b> Data Science, Bachelor of Science Research Advisor: Dr. Jingbo Shang	2017 - 2021 Cum Laude

**Publications**

[7] **Zhaoyi Hou**, Li Zhang, and Chris Callison-Burch. Choice-75: A Dataset on Decision Branching in Script Learning [LREC-COLING2024]

[6] Alyssa Hwang\*, Bryan Li\*, **Zhaoyi Hou\***, Dan Roth. Large Language Models as Sous Chefs: Revising Recipes with GPT-3. [arxiv]

[5] Tianyi Zhang\*, Isaac Tham\*, **Zhaoyi Hou\***, Jiaxuan Ren, Liyang Zhou, Hainiu Xu, Li Zhang, Lara J. Martin, Rotem Dror, Sha Li, Heng Ji, Martha Palmer, Susan Brown, Reece Suchocki, and Chris Callison-Burch. Human-in-the-Loop Schema Induction. [ACL2023]

[4] Xiaochen Kev Gao, **Zhaoyi Hou**, Yifei Ning, Jingbo Shang, Vish Krishnan. *Towards Comprehensive Patent Approval Predictions: Beyond Traditional Document Classification*. [ACL2022]

[3] Caitlin A. Stamatis, Jonah Meyerhoff, Tingting Liu, **Zhaoyi Hou**, Garrick Sherman, Brenda L. Curtis, Lyle H. Ungar, David C. Mohr. *The Association of Language Style Matching in Text Messages with Symptoms of Affective Psychopathologies*. [Procedia Computer Science]

[2] Artemis Panagopoulou, Manni Arora, ...(6 more), **Zhaoyi Hou**, Alyssa Hwang, Lara Martin, Sherry Shi, Chris Callison-Burch, Mark Yatskar. *QuakerBot: A Household Dialog System Powered by Large Language Models*. [Alexa Prize TaskBot Challenge Proceedings]

[1] Emily Nicole Corbett Manoogian, Adena Zadourian, ...(8 more) **Zhaoyi Hou**, Jason G. Fleischer, Shah Golshan, Pam R Taub, Satchidananda Panda. *Feasibility of Time-Restricted Eating and Impacts on Cardiometabolic Health in 24-Hour Shift Workers: The Healthy Heroes Randomized Clinical Trial*. [Cell Metabolism]

(\*Equal contribution)

**Research Experience**

**Computer and Information Science (CIS) at Penn Engineering** Sep 2022 - May 2023  
*NLP Researcher (advised by Prof. Chris Callison-Burch)*  
- Building a human-in-the-loop script curation system to validate and improve event-based scripts generated by pre-trained large language models (e.g. GPT-3).

**Computer and Information Science (CIS) at Penn Engineering** May 2022 - May 2023  
*NLP Researcher (advised by Li Zhang and Prof. Chris Callison-Burch)*  
- Building a natural language generation model to generate branching options in daily scripts;  
- Building an explanation generation module to generate rationales for each option based on common sense.

**Computer and Information Science (CIS) at Penn Engineering** Jan 2022 - Aug 2022  
*NLP & Psychology Researcher (advised by Prof. Lyle Ungar)*  
- Defined and built a pipeline to extract semantic mirroring from mobile text conversations;  
- Analyzed the correlation between psychological conditions and semantic mirroring in messages.

**Shang Data Lab at UCSD (SDLab)** Jun 2020 - Apr 2021  
*NLP Researcher (advised by Prof. Jingbo Shang)*  
- Built a text data ETL and classification pipeline to handle 600,000 patent documents;  
- Implemented a customized BERT-based text classification model and improved true negative rate (specificity) from 60% to 86% (heavily unbalanced data with 84% positive instance).

**Salk Institute for Biological Studies** Jul 2019 - Jun 2021  
*Data Science Researcher (advised by Prof. Satchidananda Panda)*  
- Built a data ETL and analysis pipeline for user behavior analysis (more than 500,000 records).

## Projects

**Amazon Alexa TaskBot Competition** Nov 2021 - Apr 2022  
*Information Retrieval*  
- Implemented the document retrieval module for the Alexa TaskBot competition;  
- Improved the retrieval success rate by 25% and advanced to the final list.

**Stack Overflow Question Quality Classification** May 2021 - Jul 2021  
*Text Mining & Cloud Computing*  
- Built a text data ETL and analysis pipeline for 60,000 Stack Overflow question texts;  
- Built an XGBoost classification model with AWS SageMaker and deployed it as an AWS SageMaker Endpoint;  
- Achieved 87% accuracy in the "High-Quality Question" classification task.

**Machine Learning for Ophthalmological Diagnosis** Nov 2019 - Dec 2020  
*Computer Vision & Healthcare*  
- Built an image classification pipeline to pre-diagnose common eye diseases with a convolutional neural network;  
- Achieved 75% accuracy for classifying involuntional ptosis, thyroid eye disease, and normal eyes.

## Teaching

**University of Pittsburgh** Teaching Assistant  
Jan 2024 - Present  
- CS1503 (Mathematical Foundation for Machine Learning)  
- CS1671 (Human Language Technologies)

**Penn Engineering Online** Course Development Assistant  
Oct 2022 - May 2023  
- CIS5300 (Computational Linguistics)

**Halicioğlu Data Science Institute, UC San Diego** Student Tutor  
Mar 2018 - Aug 2019, Mar 2021 - Jun 2021  
- CSE151A (Intro to Machine Learning)  
- DSC20 (Intro to Data Structure)